(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10260692 A**

(43) Date of publication of application: **29 . 09 . 98**

(51) Int. Cl **G10L 3/00**
G10L 3/00
G10L 3/00

(21) Application number: **09064933**

(22) Date of filing: **18 . 03 . 97**

(71) Applicant: **TOSHIBA CORP**

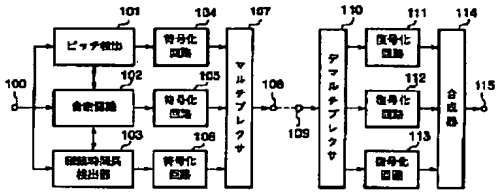
(72) Inventor: **AKAMINE MASAMI**
KOSHIBA AKINORI

(54) **METHOD AND SYSTEM FOR RECOGNITION SYNTHESIS ENCODING AND DECODING OF SPEECH**

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a speech encoding/decoding system based upon recognition synthesis which can be applied with incomplete speech recognition technology to encode a speech signal at a very low rate of 1kbps or less and transmit even nonlinguistic information on a feeling, etc., of a speaker.

SOLUTION: On a transmission side, input speech data are inputted to a pitch detection part 101, a phoneme recognition part 102, and a continuance detection part 103 to detect a pitch period, recognize a syllable, and the continuance of a phoneme, information on the pitch period, syllable, and continuance is encoded by encoding circuits 104, 105, and 106, and then the code sequence is transmitted to a channel through a multiplexer 107. On a reception side, a demultiplexer 110 decode the code sequence into the information on the pitch period, syllable, and continuance and on the basis of the decoded information, a synthesizer 114 synthesizes the original speech signal.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－260692

(43)公開日　平成10年(1998)９月29日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00		Ｇ１０Ｌ　3/00	Ｒ
			Ｈ
	５３５		５３５
	５５１		５５１Ａ

審査請求　未請求　請求項の数６　ＯＬ　（全 14 頁）

(21)出願番号　特願平9－64933

(22)出願日　平成９年(1997)３月18日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地

(72)発明者　赤嶺　政巳
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内

(72)発明者　小柴　亮典
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内

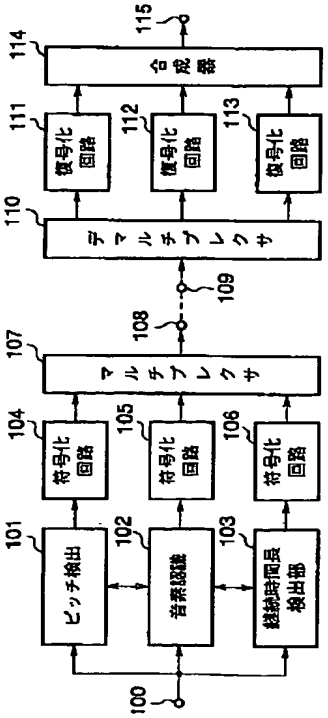
(74)代理人　弁理士　鈴江　武彦　（外６名）

(54)【発明の名称】　音声の認識合成符号化／復号化方法及び音声符号化／復号化システム

(57)【要約】

【課題】１ｋｂｐｓ以下の極低レートで音声信号を符号化するために、不完全な音声認識技術でも適用でき、かつ話者の感情など非言語的な情報も伝送することができる認識合成に基づいた音声符号化／復号化システムを提供する。

【解決手段】送信側において入力音声データをピッチ検出部１０１、音素認識部１０２および継続時間長検出部１０３に入力して、ピッチ周期の検出、音節の認識および音素の継続時間長の検出を行い、これらピッチ周期、音節および継続時間長の情報を符号化回路１０４，１０５，１０６により符号化した後、符号列をマルチプレクサ１０７を経て通信路に伝送し、受信側においてはデマルチプレクサ１１０で符号列からピッチ周期、音節および継続時間長の情報を復号化し、これらの復号化された情報に基づいて合成器１１４で元の音声信号を合成する。



1

【特許請求の範囲】

【請求項１】入力音声信号から文字情報を認識するとともに、該入力音声信号から韻律情報を検出して、これら文字情報および韻律情報を符号化データとして伝送または蓄積し、伝送または蓄積された符号化データから前記文字情報および韻律情報を復号し、復号された文字情報および韻律情報に基づいて音声信号を合成することを特徴とする音声の認識合成符号化／復号化方法。

【請求項２】入力音声信号から音素、音節または単語を文字情報として認識するとともに、該入力音声信号からピッチ周期と前記音素または音節の継続時間長を韻律情報として検出して、これら文字情報および韻律情報を符号化データとして伝送または蓄積し、伝送または蓄積された符号化データから前記文字情報および韻律情報を復号し、復号された文字情報および韻律情報に基づいて音声信号を合成することを特徴とする音声の認識合成符号化／復号化方法。

【請求項３】入力音声信号から文字情報を認識する認識手段と、

前記入力音声信号から韻律情報を検出する検出手段と、

前記文字情報および韻律情報を符号化する符号化手段と、

前記符号化手段により得られた符号化データを伝送または蓄積する伝送／蓄積手段と、

前記伝送／蓄積手段により伝送または蓄積された符号化データから前記文字情報および韻律情報を復号する復号化手段と、

前記復号化手段により復号された文字情報および韻律情報に基づいて音声信号を合成する合成手段とを備えたことを特徴とする音声符号化／復号化システム。

【請求項４】入力音声信号から音素、音節または単語を文字情報として認識する認識手段と、

前記認識手段により認識された文字情報の継続時間長を検出する継続時間長検出手段と、

前記入力音声信号のピッチ周期を検出するピッチ検出手段と、

前記文字情報と、前記継続時間長およびピッチ周期からなる韻律情報を符号化する符号化手段と、

前記符号化手段により得られた符号化データを伝送または蓄積する伝送／蓄積手段と、

前記伝送／蓄積手段により伝送または蓄積された符号化データから前記文字情報および韻律情報を復号する復号化手段と、

前記復号化手段により復号された文字情報および韻律情報に基づいて音声信号を合成する合成手段とを備えたことを特徴とする音声符号化／復号化システム。

【請求項５】前記合成手段は、前記音声信号の合成に用いる合成単位の情報を格納した合成単位辞書として、異なる話者の音声データから生成された複数個の合成単位辞書を備え、前記韻律情報に応じて該複数個の合成単位辞書の中から１個の合成単位辞書を選択して前記音声信号を合成することを特徴とする請求項３または４記載の音声符号化／復号化システム。

2

【請求項６】前記合成手段は、前記音声信号の合成に用いる合成単位の情報を格納した合成単位辞書として、異なる話者の音声データから生成された複数個の合成単位辞書を備え、指示された合成音の種類に応じて該複数個の合成単位辞書の中から１個の合成単位辞書を選択して前記音声信号を合成することを特徴とする請求項３または４記載の音声符号化／復号化システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、音声信号を高能率に圧縮符号化／復号化する方法及びシステムに係り、特に音声信号を１ｋｂｐｓ以下の極低ビットレートで符号化する認識合成符号化方法及びこれを用いた音声符号化／復号化方法及びシステムに関する。

【０００２】

【従来の技術】音声信号を高能率に符号化する技術は、利用できる電波帯域が限られている移動体通信や、メモリの有効利用が求められるボイスメールなどの蓄積媒体において、今や不可欠の技術になっており、より低いビットレートへ向かっている。電話帯域の音声を４ｋｂｐｓ～８ｋｂｐｓ程度の伝送レートで符号化する方式として、ＣＥＬＰ(Code Excited Linear Prediction)は有効な方式の一つである。

【０００３】このＣＥＬＰ方式に関しては、M.R.Schroeder and B.S.Atal, “Code ExcitedLinear Prediction (CELP):High Quality Speech at Very Low Bit Rates”, Proc. ICASSP, pp. 937-940, 1985およびW.S.Kleijin, D.J.Krasinski et al. “Improved Speech Quality and Efficient Vector Quantization in SELP”, Proc. ICASSP, pp. 155-158, 1988（文献１）で詳しく述べられている。

【０００４】同文献１によると、この方式はフレーム単位に分割された入力音声から、声道をモデル化した音声合成フィルタを求める処理と、このフィルタの入力信号に当たる駆動ベクトルを求める処理に大別される。これらのうち、後者は符号帳に格納された複数の駆動ベクトルを一つずつ音声合成フィルタに通し、合成音声と入力音声との歪を計算し、この歪が最小となる駆動ベクトルを探索する処理からなる。これは閉ループ探索と呼ばれており、４ｋｂｐｓ～８ｋｂｐｓ程度のビットレートで良好な音質を再生するために非常に有効な方法である。

【０００５】また、音声信号を更に低いビットレートで符号化する方法として、ＬＰＣボコーダが知られている。これは声帯信号をパルス列と白色雑音信号で、また声道の特性をＬＰＣ合成フィルタでモデル化し、それらのパラメータを符号化する方式であり、音質的に問題はあるものの音声信号を２．４ｋｂｐｓ程度で符号化する

ことができる。これらの符号化方式は、発声者が何を言っているかという言語情報はもちろん、個人性、声の質、感情など元の音声波形が持っている情報を人間の聴覚特性上できるだけ忠実に伝送しようとするもので、主に電話を中心とする通信の用途に用いられている。

【０００６】一方、最近のインターネットブームを背景にネットチャットと呼ばれるサービスの利用者が増加している。これは、ネットワーク上でリアルタイムに一対一、または一対多、多対多の会話を楽しむものであり、音声信号の伝送のため上記のＣＥＬＰ方式を基本にしたものが用いられている。ＣＥＬＰ方式は、ＰＣＭ方式と比べビットレートが１／８～１／１６と低く、音声信号の能率的な伝送を可能にしている。しかし、インターネットを利用するユーザ数は急激に増加しつつあり、これに伴いネットワークがしばしば混雑する状況が発生し、そのため音声情報の伝送に遅延が生じて会話に支障が起こっている。

【０００７】このような状況を解決するためには、音声信号をＣＥＬＰ方式よりさらに低いビットレートで符号化する技術が必要である。低ビットレート符号化の究極の姿としては、音声の言語情報を認識してその言語情報を表現する文字列を伝送し、受信側で規則合成する認識合成符号化が知られている。この認識合成符号化は、中田和男著、「音声の高能率符号化」、森北出版発行（文献２）で簡単に紹介されているように、数十乃至１００ｂｐｓ程度の極低レートで音声信号を伝送することができると言われている。

【０００８】しかし、認識合成符号化方式は音声認識技術を適用することで得られた文字列から音声を規則的に合成する必要があるため、音声認識が不完全であるとイントネーションが著しく不自然になったり、会話の内容が誤ったりという問題が生じる。このため、認識合成符号化は完全な音声認識技術を仮定しており、今まで具体的に実現された例はなく、近い将来もその実現は困難であると予想される。

【０００９】このように音声信号という物理的な情報を言語情報という高度に抽象化された情報に変換した後、通信を行う方法では、実現性に問題があるため、音声信号をより物理的な情報に認識して変換する符号化方法が提案されている。この方法の一例として、特公平５－７６０４０号（文献３）に記述されている「ボコーダ方法及び装置」が知られている。

【００１０】同文献３においては、アナログ音声入力は音声認識装置へ送られ、音素列に変換される。音素列は、音素－異音合成器によってそれを近似した異音列に変換され、この異音列によって音声が再生される。音声認識装置では、アナログ音声入力はＡＧＣにより信号のゲインがある一定の値に保持されつつ、ホルマントトラッカーに入力されて入力信号のホルマントが検出され、ＲＡＭに記憶される。記憶されたホルマントは音素境界検出装置へ送られ、音素の成分へ区切られる。区切られた音素は、認識アルゴリズムによって音素テンプレート登録表との間でマッチングがとられ、認識された音素が得られる。

【００１１】音素－異音合成器では、入力された音素符号と対応する異音列をＲＯＭから読み出し、音声合成器へ送る。音声合成器は送られてきた異音列から線形予測フィルタのパラメータなど音声合成に必要なパラメータを求め、それらを用いて音声を合成する。ここで、異音（Ａｌｌｏｐｈｏｎｅ）と呼ぶものは、当該音素とその前後の音素を用いて予め定めた規則に従って決定された属性（当該音素が語頭、語中、語尾のどれか、また、鼻音化や無声化するかといった類）が付加されたものである。

【００１２】この文献３に記載された技術のポイントは、音声信号を言語情報である文字列に変換するのではなく、音声信号を単に音素記号列に変換し、その記号列と音声合成のための物理パラメータを対応付けたことである。このようにすることによって、音素の認識が誤ったとしても、誤った音素が別の音素に変わるものの文章全体としては大きく変わらないという利点が生じる。そして、文献３には「人間の耳の自然のフィルタ作用と聞き手の思考過程での誤り修正のために、完全な認識でなくても、最も良い一致を取ることで、認識アルゴリズムによって発生する誤りは最小のものとなる。」と記載されている。

【００１３】しかし、文献３に記載の符号化方法では、符号化側から単に音素を表す記号列を伝送しているのみであるため、復号化側で再生される合成音は抑揚やリズムのない不自然なものとなってしまい、単に会話の意味が伝わるのみで話者に関するの情報や感情といった情報は伝わらないという問題がある。

【００１４】

【発明が解決しようとする課題】上述したように、音声の言語情報を認識し、その情報を表現する文字列を伝送、復号化側で規則合成する従来の認識合成方式は完全な音声認識技術を仮定しているため、現実に実現することが困難であるという問題があった。

【００１５】また、不完全な音声認識技術でも適用できる公知の符号化方式では、単に音素を表す記号列を伝送しているのみであるため、復号化側で再生される合成音は抑揚やリズムのない不自然なものとなってしまい、単に会話の意味が伝わるのみで話者に関するの情報や感情といった情報は伝わらないという問題があった。

【００１６】本発明は、１ｋｂｐｓ以下の極低レートで音声信号を符号化するために、不完全な音声認識技術でも適用でき、かつ話者の感情など非言語的な情報も伝送することができる認識合成に基づいた音声の認識合成符号化／復号化方法及びシステムを提供するものである。

【００１７】

5

【課題を解決するための手段】上記の課題を解決するため、本発明に係る音声の認識合成符号化／復号化方法は、入力音声信号から音素、音節または単語などを文字情報として認識するとともに、該入力音声信号からピッチ周期と音素または音節の継続時間長などを韻律情報を検出して、これら文字情報および韻律情報を符号化データとして伝送または蓄積し、伝送または蓄積された符号化データから文字情報および韻律情報を復号し、復号された文字情報および韻律情報に基づいて音声信号を合成することを特徴とする。

【００１８】また、本発明に係る音声符号化／復号化システムは、入力音声信号から文字情報を認識する認識手段と、入力音声信号から韻律情報を検出する検出手段と、これら文字情報および韻律情報を符号化する符号化手段と、この符号化手段により得られた符号化データを伝送または蓄積する伝送／蓄積手段と、この伝送／蓄積手段により伝送または蓄積された符号化データから文字情報および韻律情報を復号する復号化手段と、この復号化手段により復号された文字情報および韻律情報に基づいて音声信号を合成する合成手段とを備えたことを特徴とする。

【００１９】より具体的には、認識手段は入力音声信号から音素、音節または単語を文字情報として認識し、韻律情報検出手段は認識された文字情報の継続時間長と入力音声信号のピッチ周期を韻律情報として検出する。

【００２０】このように本発明では、符号化側（送信側）において入力音声信号から音素や音節などの文字情報を認識してその情報を伝送または蓄積することに加えて、入力音声信号からピッチ周期や継続時間長などの韻律情報を検出してその情報も伝送または蓄積し、復号化が（受信側）において伝送または蓄積されてきた音素や音節などの文字情報とピッチ周期や継続時間長などの韻律情報に基づいて音声信号を合成することにより、１ｋｂｐｓ以下といった極低レートでの符号化が可能である上に、話者の抑揚やリズム、話調なども再生されることによって、従来では困難であった話者の感情などの非言語的情報の伝送も可能となる。

【００２１】また、本発明においては音声信号の合成に用いる合成単位の情報を格納した合成単位辞書として異なる話者の音声データから生成された複数個の合成単位辞書を備え、韻律情報に応じて１個の合成単位辞書を選択して音声信号を合成するようにしてもよい。このように構成にすると、符号化側（送信側）で音声信号を入力した話者とより類似した合成音が復号化側（受信側）で再生される。

【００２２】さらに、上述した複数個の合成単位辞書の中から、指示された合成音の種類に応じて１個の合成単位辞書を選択して音声信号を合成するようにしてもよい。このようにすると、合成される音声信号の種類を送信側または受信側のユーザを指定でき、声質変換なども

6

可能となる。

【００２３】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を説明する。

（第１の実施形態）図１は、本発明の第１の実施形態に係る音声の認識合成符号化／復号化方法を適用した音声符号化／復号化システムの構成を示すブロック図である。この符号化／復号化システムは、ピッチ検出部１０１、音素認識部１０２、継続時間長検出部１０３、符号化回路１０４，１０５，１０６、マルチプレクサ１０７、デマルチプレクサ１１０、復号化回路１１１，１１２，１１３および合成器１１４から構成される。

【００２４】まず、符号化側（送信側）においては、音声入力端子１００からディジタル化された音声信号（以下、入力音声データという）が入力される。この入力音声データはピッチ検出部１０１、音素認識部１０２、継続時間長検出部１０３に入力される。ピッチ検出部１０１による検出結果、音素認識部１０２による認識結果および継続時間長検出部１０３による検出結果は、符号化回路１０４，１０５，１０６によってそれぞれ符号化された後、符号化多重化部であるマルチプレクサ１０７により多重化されて符号化列となり、出力端子１０８から通信路に伝送される。

【００２５】一方、復号化側（受信側）においては、符号化側（送信側）から通信路を経て伝送されてきた符号列が符号化分解部であるデマルチプレクサ１１０によって、ピッチ周期の符号、音素の符号、継続時間長の符号に分解された後、復号化回路１１１，１１２，１１３に入力されて元のデータが復号され、さらに合成器１１４により合成されて出力端子１１５から合成音声信号（復号音声信号）が出力される。

【００２６】次に、図１の各部について詳細に説明する。音素認識部１０２は、公知の認識アルゴリズムを用いて音声入力端子１００からの入力音声データに含まれる文字情報を音素単位で識別し、その識別結果を符号化回路１０４に出力する。認識のアルゴリズムとしては、北脇信彦編著、「音のコミュニケーション工学」コロナ社などのテキストで紹介されているように、種々の方法を用いることができる。ここでは、音素を認識単位とするアルゴリズムとして以下の方法を用いるものとする。

【００２７】図２に、このアルゴリズムに基づく音素認識部１０２の構成を示す。この音素認識部１０２において、音声入力端子１００からの入力音声データはまず分析フレーム生成部２０１に入力される。分析フレーム生成部２０１は、入力音声データを分析フレーム長に分割し、さらに窓関数をかけて信号の打ち切りによる影響を減じてから、結果を特徴量抽出部２０２に送る。特徴量抽出部２０２は、分析フレーム毎にＬＰＣケプストラム係数を計算し、これを特徴ベクトルとして音素判定部２０３に送る。音素判定部２０３は、入力された分析フレ

ーム毎の特徴ベクトルと、代表特徴量メモリ２０４に予め用意してある音素毎の代表的な特徴ベクトルとの間の類似度としてユークリッド距離を計算し、この距離が最も小さい音素をそのフレームの音素と判定し、この判定結果を出力する。

【００２８】ここでは特徴量としてＬＰＣケプストラム係数を用いたが、これにさらにΔケプストラムを併せて用いることにより、認識率を向上させることも可能である。また、入力された分析フレームのＬＰＣケプストラム係数だけを特徴ベクトルとするのではなく、そのフレームの前後一定時間に入力された分析フレームから得られるＬＰＣケプストラム係数も含めて特徴ベクトルとすることにより、ＬＰＣケプストラム係数の時間方向の変動を考慮する方法もある。さらに、ここでは特徴ベクトルの間の類似度としてユークリッド距離を用いたが、特徴ベクトルにＬＰＣケプストラム係数を用いていることを考慮して、ＬＰＣケプストラム距離を用いることもできる。

【００２９】ピッチ検出部１０１は、音素認識部１０２の動作と同期をとりながら、または予め定めた単位時間毎に、音声入力端子１００からの入力音声データが有声音か無声音かの判定を行い、有声音と判定した場合には更にピッチ周期を検出する。ここで得られた有声音／無声音判定結果とピッチ周期の情報は符号化回路１０５に送られ、有声音／無声音判定結果とピッチ周期を表す符号が割り当てられる。有声音／無声音判定とピッチ周期検出のアルゴリズムとしては、自己相関法など既知の手法を用いることができる。この場合、音素認識部１０２の認識結果とピッチ検出部１０１の検出結果を互いに利用し合うことで、音素認識、ピッチ検出の精度を改善させることができる。

【００３０】継続時間長検出部１０３は、音素認識部１０２の動作と同期をとりながら、音素認識部１０２で認識された音素の継続時間長を検出する。図３に示すフローチャートを参照して、継続時間長の検出手順の一例を説明する。

【００３１】まず、ステップＳ１１で音素認識を実行するための分析フレーム長を設定し、ステップＳ１２で音素認識を実行するフレーム番号を初期化する。次に、ステップＳ１３で音素の認識を音素認識部１０２により実行し、ステップＳ１４でその認識結果が前フレームの認識結果と同じかどうか判定する。現フレームと前フレームの音素の認識結果が同じである場合は、ステップＳ１５でフレーム番号をインクリメントしてステップＳ１３に戻り、そうでない場合はステップＳ１６でフレーム番号ｎを出力する。以上の処理を入力音声データがなくなるまで行う。

【００３２】こうして検出される音素の継続時間時間長は、ｎとフレーム長の積になる。また継続時間長の検出に関しては、ある音素が認識されたとき、次に別の音素が認識されるまでに最低要する時間を予め定めておき、音素の認識誤りによって、実際にはありえない継続時間長が出力されることを抑制する方法も考えられる。継続時間長検出部１０３の検出結果は符号化回路１０６に送られ、継続時間長を表す符号が割り当てられる。

【００３３】符号化回路１０４，１０５，１０６の出力は符号多重化部１０７に送られ、ピッチ周期の符号、音素の符号および継続時間長の符号が多重化されて符号列となり、出力端子１０８から通信路に伝送される。以上が符号化側（送信側）の動作である。

【００３４】復号化側（受信側）では、入力端子１０９から入力された符号列がまず符号分解部１１０でピッチ周期の符号と音素の符号、継続時間長の符号に分解され、それぞれ復号化回路１１１，１１２，１１３に出力される。復号化回路１１１，１１２，１１３では、それぞれピッチ周期、音素、継続時間長が元のデータに復号され、それらのデータが合成器１１４に送られる。合成器１１４はピッチ周期、音素、継続時間長のデータを用いて音声信号を合成する。

【００３５】合成器１１４での合成方式としては、北脇信彦編著、「音のコミュニケーション工学」コロナ社で紹介されているように、合成単位の選択と合成に用いるパラメータの選択の組合せによって種々の方式を用いることができる。ここでは、音素を合成単位とする方式の例として、特公昭５９－１４７５２に開示されている分析合成方式による合成器を用いるものとする。

【００３６】図４に、この方式による合成器１１４の構成を示す。まず、入力端子３００，３０１，３０２からピッチ周期、音素、継続時間長のデータが入力され、これらが入力バッファ３０３に書き込まれる。パラメータ結合処理部３０５は、入力バッファ３０３から音素のデータ系列を読み出し、各音素に対応したスペクトルパラメータをスペクトルパラメータメモリ３０４から読み出して単語あるいは文として結合し、バッファ３０７に出力する。スペクトルパラメータメモリ３０４には、合成単位である音素がＰＡＲＣＯＲ、ＬＳＰ、ホルマントなどのスペクトルパラメータの形で表現され、予め蓄積されている。

【００３７】音源生成処理部３０６は、入力バッファ３０３から音素、ピッチ周期、継続時間長のデータ系列を読み出し、これらのデータに基づいて音源波形メモリ３１１から音源波形を読み出し、ピッチ周期と継続時間長に基づいて、この音源波形を加工することにより、合成フィルタ３０９の駆動音源信号を生成する。音源波形メモリ３１１には、実音声データ中の各音素信号を線形予測分析して得られる残差信号から抽出された音源波形が蓄積されている。

【００３８】音源生成処理部３０６での駆動音源信号の生成は、合成する音素が有声音のときと無声音のときで処理が異なる。合成する音素が有声音のときは、音源波

形を入力バッファ３０３から読み込んだ継続時間と等しい長さになるまで、入力バッファ３０３から読み込んだピッチ周期単位で重ね合せ補間または間引き処理を行うことによって、駆動音源信号が生成される。合成する音素が無声音のときは、音源波形メモリから読み出された音源波形をそのまま、または、一部を切り出したり繰り返したりして、入力バッファ３０３から読み込んだ継続時間と等しい長さに加工することにより生成される。

【００３９】最後に、合成フィルタ３０９によりバッファ２０７に書き込まれたスペクトルパラメータとバッファ３０８に書き込まれた駆動音源信号が読み出され、音声合成のモデルに基づいて音声信号が合成されて合成音声信号が出力端子３１０から図１の出力端子１１５へと出力される。

【００４０】（第２の実施形態）図５に、本発明の第２の実施形態に係る音声の認識合成符号化／復号化方法を適用した音声符号化／復号化システムの構成を示す。第１の実施形態では、入力音声データの音素を認識し、合成単位を音素とする構成を示したが、第２の実施形態は合成単位を音節単位とするものである。

【００４１】図５の構成は、音節認識部１２２と合成器１２４を除いて図１の構成と基本的に同じである。合成する音節の単位や音節認識法には種々あるが、ここでは一例として合成単位をＣＶ，ＶＣ音節とし、音節認識法として以下の方法を用いる。ただし、Ｃは子音、Ｖは母音を表す。

【００４２】図６に、ＣＶ，ＶＣ音節を単位とする音節認識部１２２の構成を示す。音素認識部４０１は、前記の音素単位の認識部１０２と同じ働きをするものであり、音声信号を入力すると、フレーム毎に認識した音素を出力する。ＣＶ音節を単位とする音節認識部４０２は音素認識部４０１から出力された音素列からＣＶ音節を認識して出力する。ＶＣ音節構成部４０３はＣＶ音節認識部４０２から出力されたＣＶ音節列からＶＣ音節を構成し、これを入力と合わせて結果を出力する。

【００４３】図７のフローチャートを参照して、ＣＶ音節認識部４０２による音節認識処理手順の一例を説明する。まず、ステップＳ２１で入力音声データの先頭の音素にフラグを立てる。ステップＳ２２では、音節認識部４０１に入力する音素数ｎを予め定めておいた数Ｉに初期化する。ステップＳ２３で、実際にｎ個の連続する音素を予めＣＶ音節毎に用意した音素を出力シンボルとする離散型HMMに入力する。ステップＳ２４では、各HMM毎に、入力した音素列がそのHMMから出力される確率ｐを求める。ステップＳ２５では、ｎが予め定めておいた入力音素数の上限Ｎに達したかどうか判定する。ｎがＮに達していなければ、ステップＳ２６で入力する音素数ｎをｎ＝ｎ＋１として、ステップＳ２３から繰り返す。ｎがＮに達していれば、ステップＳ２７に進む。ステップＳ２７では、まず確率ｐを最大とするHMMに対応するＣＶ音節、および音素数ｎを求める。次に、フラグを立てた音素に対応するフレームから数えて、求めた音素数分の区間が該ＣＶ音節に対応する区間であると判定し、これを求めたＣＶ音節とともに出力する。ステップＳ２８では、音素の入力が終了したがどうか判定し、終了していない場合にはステップＳ２９で出力した区間の次の音素にフラグを立ててステップＳ２２に戻り、再びこの操作を繰り返す。

【００４４】次に、ＶＣ音節構成部４０３について説明する。ＶＣ音節構成部４０３には、前記の方法で出力されたＣＶ音節およびその音節の対応する区間が入力される。ＶＣ音節構成部４０３は、予め２つのＣＶ音節からＶＣ音節を構成するための方法を記述したメモリを有し、入力される音節列をそのメモリに従ってＶＣ音節列に再構成する。２つのＣＶ音節からＶＣ音節を構成する方法としては、１つ目のＣＶ音節の中心フレームから次のフレームの中心フレームまでの区間を１つ目のＣＶ音節の母音と次のＣＶ音節の子音からなるＶＣ音声と定めるという方法などが考えられる。

【００４５】音節を合成単位とする合成器の他の例として、特公昭５８－１３４６９７に開示された波形編集型音声合成装置を用いることができる。図８に、このような合成器１２４の構成を示す。

【００４６】図８において、制御回路５１０は入力端子５００，５０１，５０２を介してピッチ周期、音節、継続時間長のデータ系列を入力し、単位音声波形メモリ５０３に対して音節データと該メモリ５０３に蓄積されている単位音声波形の転送先を指示すると共に、ピッチ周期をピッチ変換回路５０４に送り、継続時間長を波形編集回路５０５に送る。そして、制御回路５１０は合成しようとする当該音節が有声部でピッチを変換する必要がある場合はピッチ変換回路５０４に転送し、当該音節が無声部である場合は波形編集回路５０５に転送するよう指示する。

【００４７】単位音声波形メモリ５０３は、実音声データから切り出された合成単位の音節ＣＶ，ＶＣの音声波形を蓄積しており、制御回路５１０から入力した音節データと指示に従って該当する単位音声波形をピッチ変換回路５０４または波形編集回路５０５に出力する。制御回路５１０は、ピッチを変換する必要がある場合はピッチ変換回路５０４にピッチ周期を送り、そこでピッチ周期が変換される。ピッチ周期の変換は波形重畳法など公知の方法で行われる。

【００４８】波形編集回路５０５は、制御回路５１０の指示に従ってピッチを変換する必要がある場合には、ピッチ変換回路５０４から送られた音声波形を補間または間引き処理し、また変換する必要がない場合には、単位音声波形メモリ５０３から送られた音声波形を補間または間引きすることにより入力した継続時間長と等しくなるよう処理し、音節単位の音声波形を生成する。さら

に、波形編集回路５０５は各音節の音声波形を結合することにより音声信号を作成する。

【００４９】このように図８の合成器１２４では、音節単位で音声信号を認識して合成するため、音素単位で認識して合成を行う図４に示した合成器１１４と比べて、より高音質の合成音が得られる利点がある。すなわち、音素を合成単位とする場合には、合成単位間での接続個所が多く、しかも子音から母音へ接続するように音声パラメータの変化が激しい場所でも合成単位を接続するため、高い品質の合成音を得ることが難しいのに対し、音節単位では合成単位間の接続個所が少ないばかりでなく、子音と母音の変化部を合成単位が含むため高品質の合成音が得られる。また、認識の単位が長くなることによって認識率も改善し、合成音の音質が向上する効果もある。

【００５０】（第３の実施形態）図８の合成器１２４の上述した利点に着目して、音質向上のため合成単位を音節より更に長い単語単位とすることも考えられる。しかし、合成単位が単語レベルまでになると単語を識別するための符号量が増加し、ビットレートが高くなる問題が生じる。符号量を抑えつつ、認識率を改善し音質向上を図る方法として、入力音声データを単語単位で認識し音節単位で合成する折衷案が考えられる。

【００５１】図９は、この方法に基づく本発明の第３の実施形態に係る音声符号化／復号化システムのブロック図であり、図１における音素認識部１０２または図５における音節認識部１２２が単語認識部１３２と認識された単語を音節に変換する単語－音節変換部１３３に置き換えられている点が第１および第２の実施形態と異なっている。このような構成により、符号量を増大させることなく、認識率を改善して音質の向上を図ることができる。

【００５２】（第４の実施形態）以上説明した第１、第２、第３の実施形態は、ピッチ周期や継続時間長の韻律情報を入力音声データから抽出して伝送しているものの、合成器で用いるスペクトルパラメータや音源波形、または単位音声波形は、予め作成されたある一種類のものを用いる構成となっている。このため、イントネーションやリズム、話調などの話者の韻律は復号化側で再生されるものの、再生される声の質は予め作成されたスペクトルパラメータや音源波形、または単位音声波形で定まるものとなり、話者によらず常に同一の声質が再生されてしまう。より豊かなコミュニケーションのためには、多様な声質を再生できるものが望まれる。

【００５３】本実施形態は、この要求に応えるために合成器で用いる合成単位辞書を複数備えたものである。ここで、スペクトルパラメータや音源波形、または単位音声波形などを合成単位辞書と呼んでいる。

【００５４】図１０は、本実施形態に係る成単位辞書を複数備えた符号化／復号化システムの構成を示すブロック図である。本実施形態の基本的な構成は図１、図５、図９で説明した第１、第２、第３の実施形態と同様であり、これらの実施形態と異なる点は、復号化側に複数個（Ｎ個）の合成単位辞書１４３，１４４，１４５を備え、伝送されてきたピッチ周期の情報に応じて、合成に用いる合成単位辞書を１個選択する構成としたことである。

【００５５】図１０において、符号化側の文字情報認識部１４０は、図１中に示した音素認識部１０２、図５中に示した音節認識部１２２、または図９中に示した単語認識部１３２および単語－音節変換部１３３のいずれかに相当するものである。

【００５６】一方、復号化側の復号化回路１１１は伝送されてきたピッチ周期を復号し、これを韻律情報抽出部１４１に送る。韻律情報抽出部１４１は入力されたピッチ周期を蓄積し、蓄積されたピッチ周期の系列から平均ピッチ周期やピッチ周期の最大値、最小値など韻律情報を抽出する。

【００５７】合成単位辞書１４３，１４４，１４５は、各々異なる話者の音声データから作成されたスペクトルパラメータや音源波形、または単位音声波形と各々の音声データから抽出された平均ピッチ周期やピッチ周期の最大値、最小値などの韻律情報を蓄積している。

【００５８】制御回路１４２は、韻律情報抽出部１４１から平均ピッチ周期やピッチ周期の最大値、最小値など韻律情報を受け取り、これと合成単位辞書１４３，１４４，１４５に蓄積されている韻律情報との誤差を計算し、誤差が最小となる合成単位辞書を選択して合成器１１４に転送する。ここで、韻律情報の誤差は、一例として平均ピッチ周期、最大値、最小値の各々の誤差の二乗の重み付き平均を計算することで得られる。

【００５９】合成器１１４は、復号化回路１１１，１１２，１１３からピッチ周期、音素または音節、継続時間長のデータをそれぞれ受け取り、これらのデータと制御回路１４２から転送された合成単位辞書を用いて音声を合成する。

【００６０】このような構成によると、符号化側で入力された話者と類似した声の高さの合成音が復号化側で再生されることになるため、話者の識別が容易になり、より豊かなコミュニケーションが実現される効果がある。

【００６１】（第５の実施形態）図１１に、複数の合成単位辞書を備えた別の実施形態として、第５の実施形態に係る音声符号化／復号化システムの構成を示す。この実施形態は、復号化側に複数の合成単位辞書を備えるとともに、符号化側に合成音の種類を指示するための合成音指示回路を備えることを特徴とする。

【００６２】図１１において、符号化側に設けられた合成音指示回路１５０は、復号化側で用意されている合成単位辞書１４３，１４４，１４５に関する情報を話者に提示し、どの合成音を用いるか選択させ、キーボードな

どの入力装置を通して合成音の種類を指示する合成音選択情報を受け取り、マルチプレクサ１０７に送る。話者に提示する情報は、合成単位辞書作成に用いた音声データの性別、年齢、太い声、細い声といった声質の特徴を表す情報からなる。

【００６３】マルチプレクサ１０７から通信路を経て復号化側に伝送された合成音選択情報は、デマルチプレクサ１１０を介して制御回路１５２に送られる。制御回路１５２は、合成音選択情報に基づいて合成単位辞書１４３，１４４，１４５の中から合成に用いる合成単位辞書を１個選択して合成器１１４に転送すると同時に、選択された合成単位辞書に蓄積されている平均ピッチ周期やピッチ周期の最大値、最小値などの韻律情報を韻律情報変換部１５１に出力する。

【００６４】韻律情報変換部１５１は、復号化回路１１１からピッチ周期を、また制御回路１５２から合成単位辞書の韻律情報をそれぞれ受け取り、入力したピッチ周期の平均ピッチ周期、最大値、最小値などの韻律が合成単位辞書の韻律情報に近づくようにピッチ周期を変換して、その結果を合成器１１４に与える。合成器１１４は、復号化回路１１２，１１３と韻律情報変換部１５１から音素または音節、継続時間長、ピッチ周期のデータを受け取り、これらのデータと制御回路１５２から転送された合成単位辞書を用いて音声を合成する。

【００６５】このような構成にすると、符号化側のユーザである送信者の好みによって、復号化側で再生される合成音を選択することができるばかりでなく、男性の声を女性の声で再生するというように男女間の声質の変換を含む各種声質の変換を容易に実現できる従来の符号化装置にはなかった効果が生じる。このような声質の変換など多様な合成音を実現する機能は、インターネットなどで不特定の人間同士でおしゃべりをしてコミュニケーションを図る場合、会話を楽しくしたり、豊かにするのに有効である。

【００６６】（第６の実施形態）図１２に、本発明の第６の実施形態に係る符号化／復号化システムの構成を示す。図１１に示した第５の実施形態では、符号化側に合成音指示回路１５０を備える構成としたが、図１２に示すように復号化側に合成音指示回路１６０を備える構成としてもよい。このようにすると、符号化側のユーザである受信者が再生される合成音の声質などを選択することができるという利点がある。

【００６７】（第７の実施形態）図１３に、本発明の第７の実施形態に係る符号化／復号化システムの構成を示す。本実施形態は図１１に示した第５の実施形態と同様に符号化側に合成音指示回路１５０を備え、復号化側で合成音指示回路１５０からの指示に基づいて韻律情報および合成器１１４のパラメータを変換して合成音の抑揚や声質を送信者の好みに応じて変えられるようにしたことを特徴とする。

【００６８】図１３において、符号化側に設けられた合成音指示回路１５０は、送信者の指示により例えばロボットの声、アニメーションの声、宇宙人の声など予め作成された声の特徴を表す分類の中から好みの声を選択し、それを表すコードを合成音選択情報としてマルチプレクサ１０７に送る。

【００６９】マルチプレクサ１０７から通信路を経て復号化側に伝送された合成音選択情報は、デマルチプレクサ１１０を介して変換テーブル１７０に送られる。変換テーブル１７０は、符号化側で合成音指示回路１５０を介して指示されたロボットの声、アニメーションの声、宇宙人の声などの合成音の特徴に対応して合成音の抑揚を変換するための抑揚変換パラメータと声質を変換するための声質変換パラメータを予め蓄積している。そして、変換テーブル１７０はデマルチプレクサ１１０を介して入力された合成音指示回路１５０からの合成音指示情報に従って、抑揚変換パラメータおよび声質変換パラメータの情報を制御回路１５２と韻律情報変換部１７１および声質変換部１７２に送る。

【００７０】制御回路１５２は、変換テーブル１７０からの情報に基づいて合成単位辞書１４３，１４４，１４５の中から合成に用いる合成単位辞書を１個選択して合成器１１４に転送すると同時に、選択された合成単位辞書に蓄積されている平均ピッチ周期やピッチ周期の最大値、最小値などの韻律情報を韻律情報変換部１７１に出力する。

【００７１】韻律情報変換部１７１は、制御回路１５２から合成単位辞書の韻律情報を、変換テーブル１７０から抑揚変換パラメータの情報をそれぞれ受け取り、入力したピッチ周期の平均ピッチ周期、最大値、最小値などの韻律情報を変換して、その結果を合成器１１４に供給する。一方、声質変換部１７２は制御回路１５２により選択された合成単位辞書に蓄積されている音源波形、スペクトルパラメータなどを変換して合成器１１４に送る。

【００７２】図１１に示した第５の実施形態では、合成音の抑揚や声質の種類は合成単位辞書１４３，１４４，１４５の作成時に用いられた音声の種類によって事実上制限される構成となっていたが、本実施形態によると韻律情報や音源波形、スペクトルパラメータの変換規則を多様にすることにより、合成音の種類を容易により多様なものとすることができる。

【００７３】なお、図１３では合成音指示回路１５０を符号化側に設けたが、図１２と同様に復号化側に設けてもよい。以上、本発明の実施形態をいくつか説明したが、本発明の主旨は符号化側において入力音声信号から音素、音節または単語などの文字情報を認識し、それらを伝送または蓄積するとともにピッチ周期や継続時間長などの韻律情報を検出して伝送または蓄積し、復号化側において伝送または蓄積されてきた音素、音節または単

語などの文字情報と、ピッチ周期や継続時間長などの韻律情報に基づいて音声信号を合成するものであり、この主旨の範囲内で様々な変形が可能である。また、認識の手法、ピッチ検出法、継続時間長の検出法、伝送情報の符号化法、復号化法、音声合成器の方式などは、本発明の実施形態で示したもののに限定されるものではなく、公知の種々の方法、方式を適用することができる。

【００７４】

【発明の効果】以上説明したように、本発明によれば入力音声信号から音素や音節などの文字情報を認識し、それらを伝送または蓄積するのみでなく、入力音声信号からピッチ周期や継続時間長などの韻律情報を検出してそれらも伝送または蓄積し、伝送または蓄積された音素または音節などの文字情報とピッチ周期や継続時間長などの韻律情報に基づいて音声信号を合成するため、認識合成による１ｋｂｐｓ以下の極低レートでの音声信号の符号化が可能であることに加えて、話者の抑揚やリズム、話調を再生でき話者の情緒や感情を伝えることができるという従来にない優れた効果を奏する。

【００７５】また、合成に用いるスペクトルパラメータや音源波形、または単位音声波形など合成単位辞書を複数個備え、話者のピッチ情報などの韻律情報や、ユーザの指示によって合成単位辞書を選択できるようにすれば、話者の識別が容易になる効果や、ユーザが望む多様な合成音の実現、声質変換などの機能の実現によって、コミュニケーションを楽しくしたり、豊かにするという効果が得られる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

【図２】図１における音素認識部の構成例を示すブロック図

【図３】図１における継続時間長検出の処理手順を示すフローチャート

【図４】図１における合成器の構成例を示すブロック図

【図５】本発明の第２の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

【図６】図５における音節認識部の構成例を示すブロック図

【図７】図６におけるＣＶ音節認識部の処理手順を示すフローチャート

【図８】本発明で用いる合成器の他の構成例を示すブロック図

【図９】本発明の第３の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

【図１０】本発明の第４の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

【図１１】本発明の第５の実施形態に係る音声符号化／＊復号化システムの構成を示すブロック図

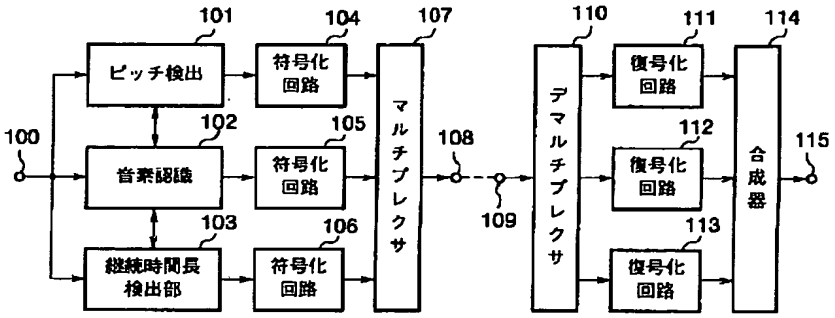
【図１２】本発明の第６の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

【図１３】本発明の第７の実施形態に係る音声符号化／復号化システムの構成を示すブロック図

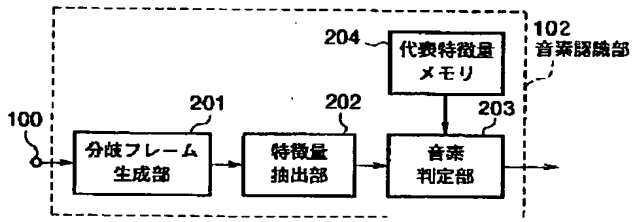
【符号の説明】

１００…音声入力端子
１０１…ピッチ検出部
１０２…音素認識部
１０３…継続時間長検出部
１０４，１０５，１０６…符号化回路
１０７…マルチプレクサ（符号多重化部）
１１０…デマルチプレクサ（符号分解部）
１１１，１１２，１１３…復号化回路
１１４…合成器
１２２…音節認識部
１３２…単語認識部
１３３…単語－音節変換部
１４０…文字情報認識部
１４１…韻律情報抽出部
１４２…制御回路
１４３，１４４，１４５…合成単位辞書
１５０…合成音指示回路
１５１…韻律情報変換部
１５２…制御回路
１６０…合成音指示回路
１７０…変換テーブル
１７１…韻律情報変換部
１７２…音質変換部
２０１…分析フレーム生成部
２０２…特徴量抽出部
２０３…音素判定部
２０４…代表特徴量メモリ
３０３…入力バッファ
３０４…スペクトルパラメータメモリ
３０５…パラメータ結合処理部
３０６…音源生成処理部
３０７，３０８…バッファ
３０９…合成フィルタ
３１１…音源波形メモリ
４０１…音素認識部
４０２…ＣＶ音節認識部
４０３…ＶＣ音節構成部
５１０…制御回路
５０３…単位音声波形メモリ
５０４…ピッチ変換回路
５０５…波形編集回路

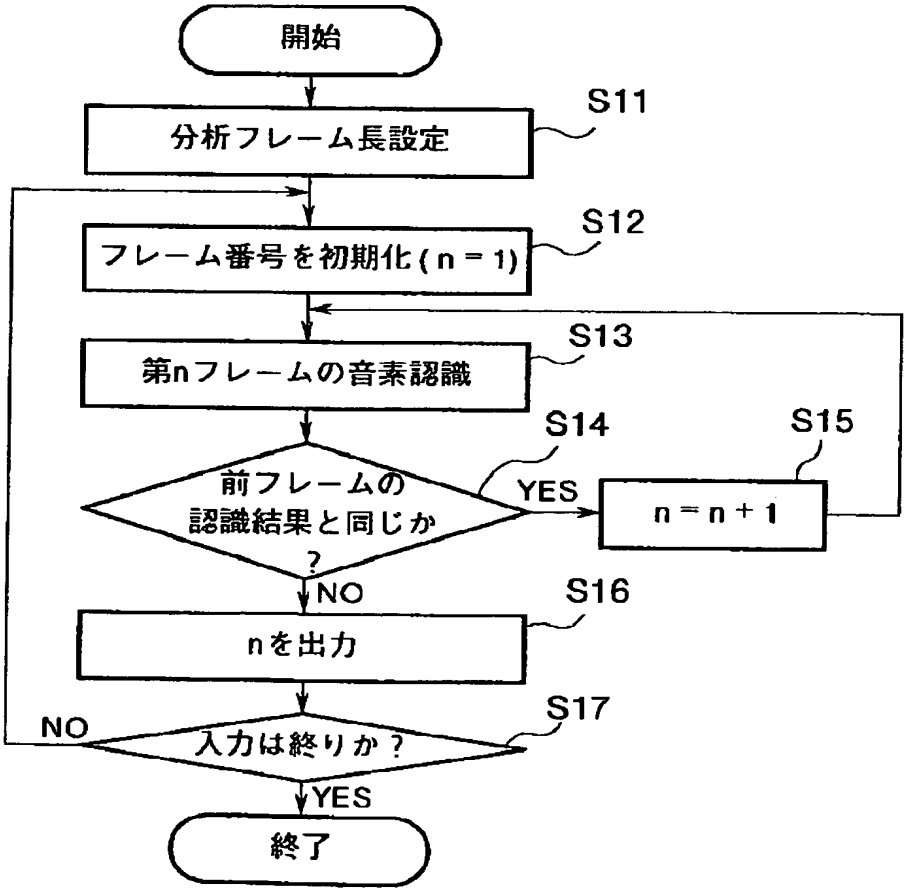
【図１】


100
101
ピッチ検出
102
音素認識
103
継続時間長検出部
104
符号化回路
105
符号化回路
106
符号化回路
107
マルチプレクサ
108
109
110
デマルチプレクサ
111
復号化回路
112
復号化回路
113
復号化回路
114
合成器
115

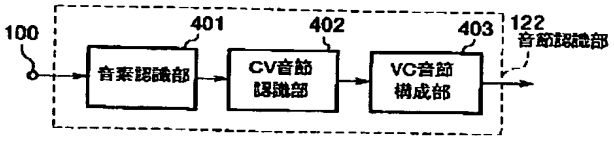
【図２】


100
201
分析フレーム生成部
202
特徴量抽出部
203
音素判定部
204
代表特徴量メモリ
102
音素認識部

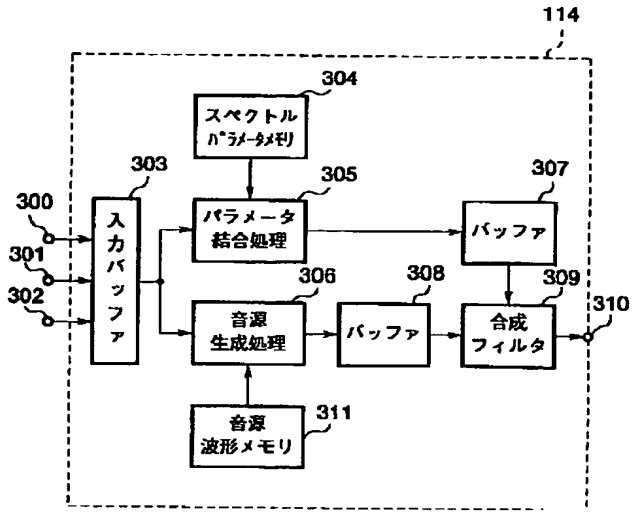
【図３】


開始
S11
分析フレーム長設定
S12
フレーム番号を初期化(n＝1)
S13
第nフレームの音素認識
S14
前フレームの認識結果と同じか？
YES
S15
n＝n＋1
NO
S16
nを出力
S17
NO
入力は終りか？
YES
終了

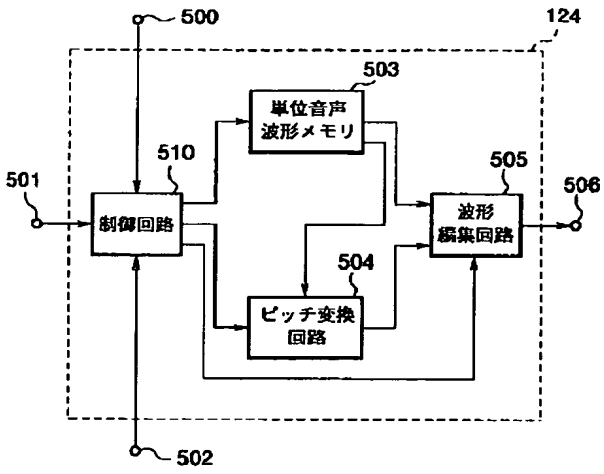
【図６】


100
401
音素認識部
402
CV音節認識部
403
VC音節構成部
122
音節認識部

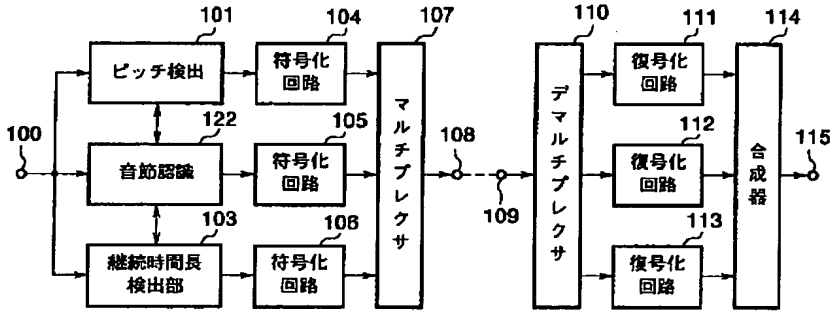
【図４】


114
304
スペクトル
ﾊﾟﾗﾒｰﾀﾒﾓﾘ
303
300
301
302
入力バッファ
305
パラメータ結合処理
307
バッファ
306
音源生成処理
308
バッファ
309
合成フィルタ
310
311
音源波形メモリ

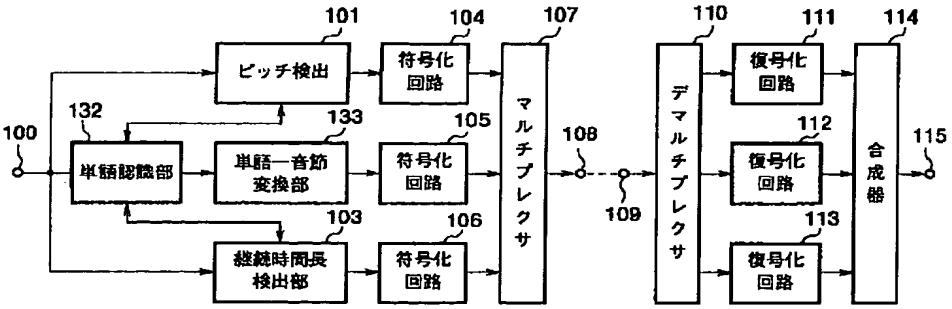
【図８】


500
124
503
単位音声波形メモリ
510
501
制御回路
505
波形編集回路
506
504
ピッチ変換回路
502

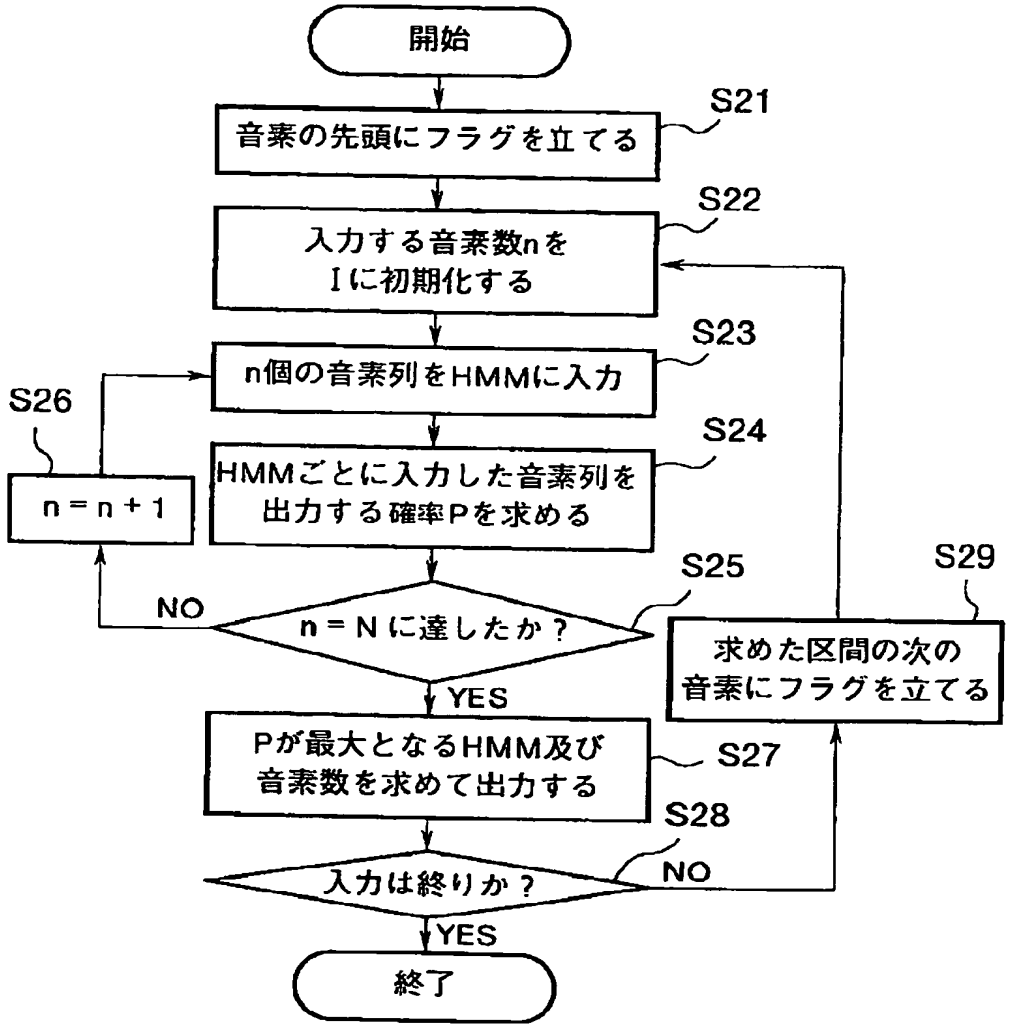
【図５】


101
104
107
110
111
114
ピッチ検出
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
100
122
105
108
109
112
115
音節認識
符号化回路
復号化回路
合成器
103
106
113
継続時間長検出部
符号化回路
復号化回路

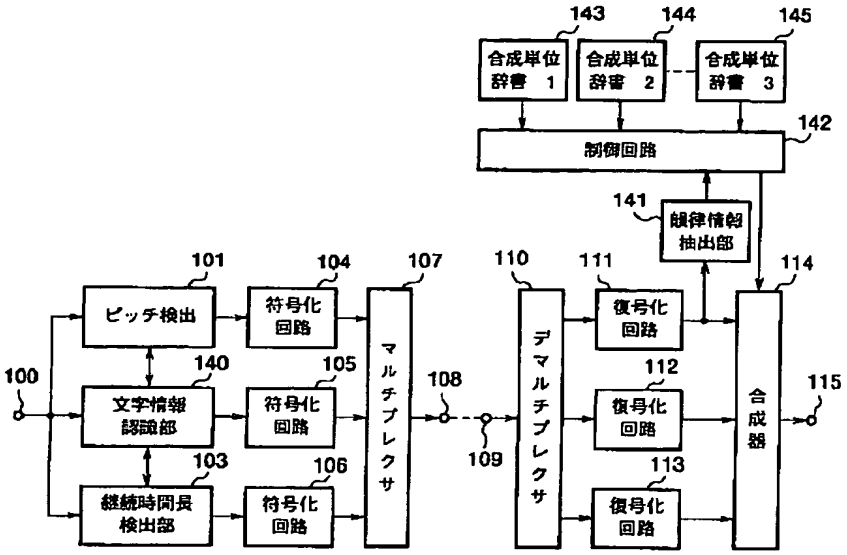
【図９】


101
104
107
110
111
114
ピッチ検出
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
132
100
133
105
108
112
115
単語認識部
単語→音節変換部
符号化回路
復号化回路
合成器
109
103
106
113
継続時間長検出部
符号化回路
復号化回路

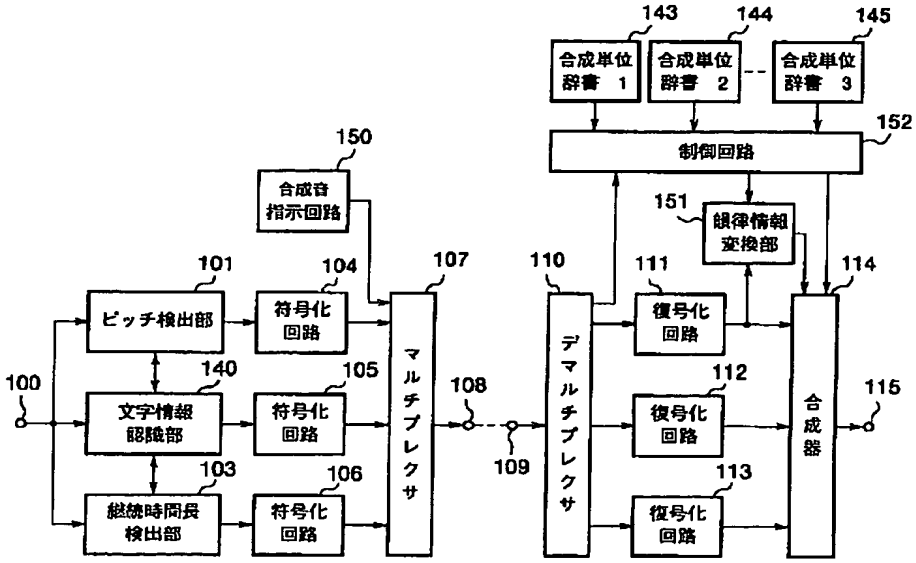
【図７】


開始
音素の先頭にフラグを立てる
S21
入力する音素数nを
Iに初期化する
S22
n個の音素列をHMMに入力
S23
HMMごとに入力した音素列を
出力する確率Pを求める
S24
n = N に達したか？
S25
NO
YES
n = n + 1
S26
求めた区間の次の
音素にフラグを立てる
S29
Pが最大となるHMM及び
音素数を求めて出力する
S27
入力は終りか？
S28
NO
YES
終了

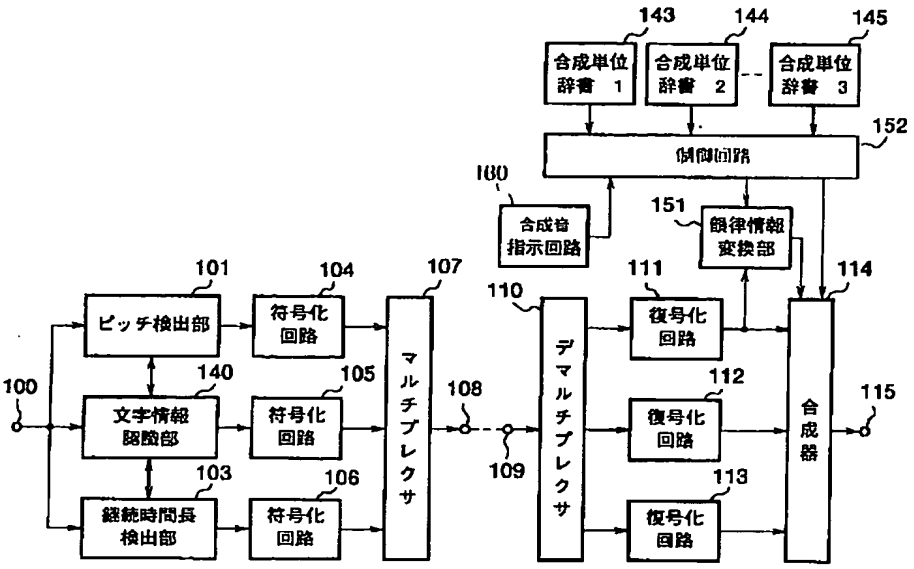
【図１０】


143
144
145
合成単位辞書 1
合成単位辞書 2
合成単位辞書 3
142
制御回路
141
韻律情報抽出部
101
104
107
110
111
114
ピッチ検出
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
合成器
100
140
105
108
112
115
文字情報認識部
符号化回路
復号化回路
109
103
106
113
継続時間長検出部
符号化回路
復号化回路

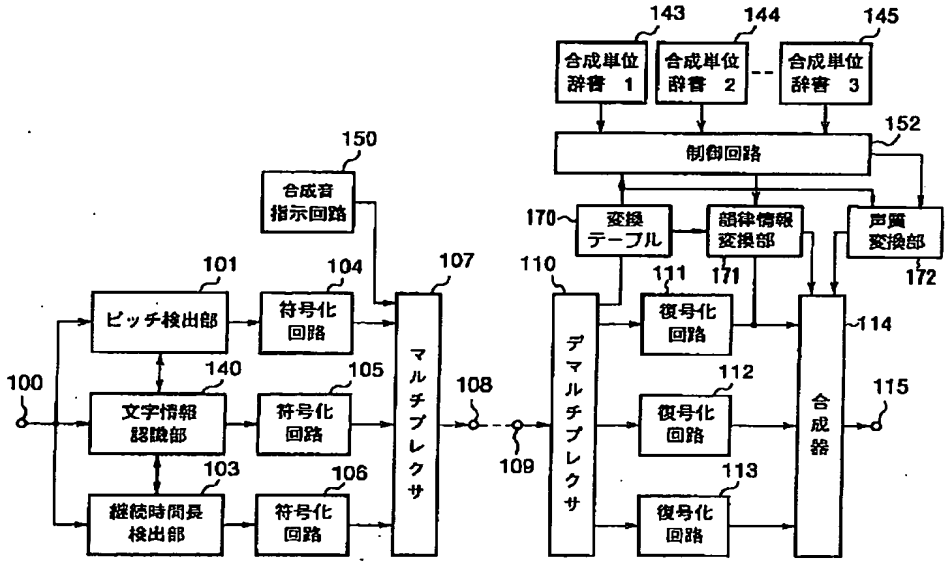
【図１１】


143
144
145
合成単位辞書 1
合成単位辞書 2
合成単位辞書 3
152
制御回路
150
合成音指示回路
151
韻律情報変換部
101
104
107
110
111
114
ピッチ検出部
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
合成器
100
140
105
108
112
115
文字情報認識部
符号化回路
復号化回路
109
103
106
113
継続時間長検出部
符号化回路
復号化回路

【図１２】


143
144
145
合成単位辞書 1
合成単位辞書 2
合成単位辞書 3
152
制御回路
180
合成音指示回路
151
韻律情報変換部
101
104
107
110
111
114
ピッチ検出部
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
合成器
100
140
105
108
112
115
文字情報認識部
符号化回路
復号化回路
109
103
106
113
継続時間長検出部
符号化回路
復号化回路

【図１３】


143
144
145
合成単位辞書 1
合成単位辞書 2
合成単位辞書 3
152
制御回路
150
合成音指示回路
170
変換テーブル
韻律情報変換部
声質変換部
172
101
104
107
110
111
171
114
ピッチ検出部
符号化回路
マルチプレクサ
デマルチプレクサ
復号化回路
合成器
100
140
105
108
112
115
文字情報認識部
符号化回路
復号化回路
109
103
106
113
継続時間長検出部
符号化回路
復号化回路